International Journal of Advanced Technology in Engineering and Science
Vol. No.7, Issue No. 03, March 2019
www.ijates.com
ijates
ISSN 2348 - 7550

AN AUTOMATIC FRAMEWORK FOR THE DETECTION OF SKIN CANCER USING DIGITAL IMAGES

Sarumathi S[1], Madhumitha S[2], Mercy S[3], Pradeeish P[4]

[1]*Department of Information Technology, K SRangasamy College of Technology, (India)*

[2]*Department of Information Technology, KSRangasamy College of Technology, (India)*

[3]*Department of Information Technology, K SRangasamy College of Technology, (India)*

[4]*Department of Information Technology, KSRangasamy College of Technology, (India)*

ABSTRACT

In today's world human cancer is the deadly disease. Among various forms of human cancer, the most common one is skin cancer. To identify skin cancer at an early stage, various techniques named as segmentation and feature extraction are used. Here we focus on the skin cancer which is the abnormal growth of skin cells. Image processing is the methodology used for the analysis and manipulation of digitalized skin cancer images. In the proposed work k-Means clustering algorithm is used for segmentation and feature extraction. Color extraction is carried out by ensemble of clusters and the creation of feature vector is implemented by using GLSM. Then, during the detection phase,the learning and prediction involve ensemble of classifiers which determines the incidence of a malignant tumour. The images are taken from the ISIC repository. The proposed system provides a skin cancer detection performance above existing system as measured by the accuracy.

Keywords:*Ensemble Classifier,Feature Extraction,Image Segmentation, Learning and Prediction*

I. INTRODUCTION

Skin cancer is a disease which involves abnormal growth of skin cells that are found in the outer layers of your skin. Our skin protects out body against heat, light, infection, and injury. It stores water, fat, and vitamin D. Skin has three main layers and several kinds of cells. Epidermis is the outer layer of the skin. Dermis is the middle one. The inner most layer of the skin is called hypodermis. The skin cancer is categorized as non-melanoma and melanoma skin cancer. Non-melanoma skin cancers involve basal cell carcinoma and squamous cell carcinoma.Melanoma is the most dangerous one. Melanoma is a cancer of melanocytes which is the pigment producing cells of the skin. It may cause death. If found at advanced stage, it's highly resistant to treatment. It affects the strangest places (including the heart and brain). Young people (those aged 25-29) are getting affected by melanoma most often. It is slightly more common in men. If it is found early, melanoma can be treated like any other skin cancer with simple treatments. The ozone layer depletion, which has diminished the protection of human skin against the UV radiation. The abusive exposure to the sun or the solarium causes skin cancer.

II.IMAGE PROCESSING

The process of converting an image into digital form and perform some operations on it is called image processing. It is used to get an enhanced image or to extract some useful information from it. Image processing involves the following three steps,

- with optical scanner used to import the image or by digital photography.
- Data compression, image enhancement and spotting patterns that are not to human eyes like satellite photographs are included for Analysing and manipulating the image
- The result can be altered based on image analysis t the output stage.

The input is given to the MATLAB in the form of images. The data set is taken from the ISIC repository. The MATLAB reads the images and analyses the images to produce the output. After, the MATLAB reads and displays the original image, it reconstructs the image, then again reconstruct the output by and then complement the result. The k-Means cluster depend on data set, it is essential to remapping the image into vector, after that determined the number of clusters, reshaping into image, and then create image segment, the last steps extracting the tumour.

The proposed system follows a pipeline of i)Image Segmentation and Feature Extraction, ii) Colour Extraction, iii) Creation of Feature Vector, iv) Learning and Prediction.

1.Segmentation and Feature Extraction

Segmentation is an important step in many perception tasks, such as some approaches to object detection and recognition (Alexander J. B. Trevor, Suat Gedikli, Radu B. Rusu, Henrik I. Christensen (2013)). Segmentation of images has been widely studied in the computer vision community, and point skin cancer segmentation has also been of interest.

For the segmentation process k-Means clustering algorithm which is one of the simplest unsupervised learning algorithms is used. This algorithm easily solves the clustering problem. The procedure gives an easier way to classify a given image through a different number of clusters k clusters. After, the MATLAB reads and displays the original image, it reconstructs the image, then again reconstruct the output by and then complement the result. The k-Means cluster depend on data set, it is essential to remapping the image into vector, after that determined the number of clusters, reshaping into image, and then create image segment, the last steps extracting the tumour.

The steps that are used the k-Means clustering are shown in Fig 1.

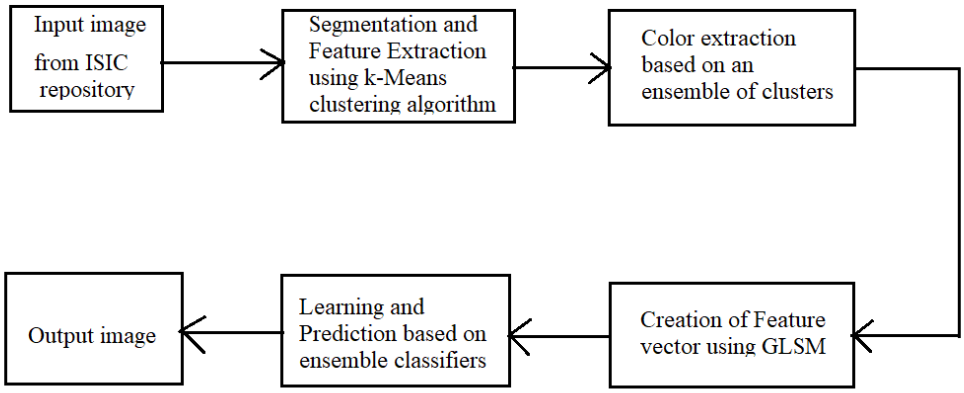

Figure 1 Block diagram of detecting skin cancer images

The feature extraction using k-Means clustering involves the following, to its nearest centre each pixel is assigned. The average location for each enrolled pixel (for a single centre) is computed once this enrolment process is completed. The new location for the specific centre is the average location. Each of the user specific centres is repeated for this process. The distance between the old location and the new location is computed.Once each of the location have been relocated,the centres no longer move or a maximum number of 1,000 iterations.

2.Color extraction based on an ensemble of clusters

For the diagnose of skin cancer, the color content of the lesion is done. Ensemble of Clusters (EoCls) is used for color segmentation to detect what colors are present in the image. EoCls are efficient because they are thought to overcome the limitations of single clustering algorithms by exploiting diversity in data processing. By using clustering algorithms on the same data or by using different values to the parameters of a single algorithm an EoCls can be obtained. The EoCls employed for detecting the RGB values of the colors that are present in a lesion is formed by three different algorithms: k-Means, and c-Means and a Kohonen map. To make the color extraction process faster, all algorithms are run in parallel, each on its own thread. Each of the clusters extracts the representative values of the partitions detected in the image being analysed. Then, a global RGB value is obtained for each color by averaging each channel representative.

3.Creation of the feature vectors

For a better classification, texture analysis is one of the most important parameters. Because they provide special characteristics present in the image. Several authors have proposed methods to extract features of images. However, those methods extract statistical properties. They do not consider both local and global

spatially correlated relationships among pixels. As opposed to the mentioned reports, we calculate the feature extraction using the GLSM method. Thefeatures vector includes: asymmetry, area of the lesion, eccentricityand the color content of the lesion.

4.Learning and prediction based on ensemble classifiers

Classification is the task of learning a target function f that maps the description of a certain set of instances to the values of a predefined attribute known as class. For solving a problem of this kind is a collection of N instances as input data, which are characterized by a tuple (X, Y) y is the attribute that indicates the class label and where X is a set of attributes[1]. Classification has two main purposes: (i) descriptive modelling that explains the behaviour between objects of different classes, and (ii) predictive modelling used for assigning a class label of an unknown instance. The classification task objective is to assign an input object x input to one of the binary output malignant tumour or benign tumour, for the problem of skin cancer detection. Input x input possesses the set of features extracted during the lesion segmentation and color extraction stages. Nonetheless, the most accurate predictions do not always provide by a single classification algorithm. An ensemble of classifiers is proposed to overcome this limitation. Ensembles of classifiers are thought to outperform individual classifiers because they allow to filter out hypothesis that are not accurate due to a small training set; help overcoming problem of local optima ensembles of classifiers are used; different classifiers expand the universe of available target functions f. The ensemble of classifiers that we developed (named MAE-NBC since it is developed following the Multi-Agent paradigm) acts on two premises: (i) the base classifiers performance and (ii) hits (H) and failures (F) obtained by base classifiers for the communication. The MAE-NBC works according to the following algorithm:

Iteration t = 0

if m > 2 are recruited and m classifier agents are started, m classifiers.
Dataset D containing features is broadcasted to classifier agent, for all i =1,…,m.
$Classifier_i$ performs a ten folds cross-validation. $F\text{-}Measure_i$ is calculated.
$Classifier_i$, for all i, i = 1…, m, constructs two subsets.

Correctly classifiedobjects contain subset H_i; subset F_i contains objects incorrectly classified.

Iteration t = 1

Aggregated sets AH and AF are formed. $AH = \cup_i H_i$; $AF = \cup_i F_i$.
$Classifier_{m+1}$, is started, based on the highest $F\text{-}Measure_i$ obtained at t=0.
$Classifier_{m+1}$ is trained with set AF. F-Measure C_{m+1} is obtained by ten folds cross validation on AF.
Classifiers 1,…,m are trained with set AH. F-Measures 1,…,m are obtained by ten folds cross-validation on AH.

Iteration t = 2

Weights are given toclassifiers1,…,m+1 according to their updated F-Measure at t=1.To reach a final conclusion weighted voting is used.

The algorithms that form the ensemble of classifier is a Naive Bayes classifier. To measure the performance of the ensemble classification metrics are obtained. We compare the performance of the MAE-NBC with those of the individual classifiers that make it up. The MAE -NBC is also contrasted with classical aggregation methods such as Bagging, Boosting and Stacking.

4.1 Classification Metrics

The following metrics are employed to evaluate the performance of classifiers: sensitivity, specificity, precision, recall, true positive rate, false positive rate, and the area under the ROC curve (AUC) and F-Measure[1]. o determine how well the segmentation algorithm performs, it requires a ground truth (GT) image, which is determined by drawing manually the border around the lesion. Using a GT image, the exclusive disjunction (XOR) operation is calculated. For digital images, sensitivity measures the proportion of actual lesion pixels that are correctly identified as such. The proportion of background skin pixels that are correctly identified by the specificity measures. Generalizing:

- TP (true positive). Correctly classified objects as the object of interest.
- FP (false positive). Incorrectly identified objects as the object of interest.
- TN (true negative). Correctly identified objects as not being the object of interest.
- FN (false negative).Incorrectly identified objects as not being the object of interest.

Sensitivity and specificity are given by:

$$sensitivity = TP/(TP + TN) \quad (1)$$

$specificity = TN/(FP + TN)$ (2)

The precision of a classifier is the fraction of tuples that were correctly classified as positive from all the tuples that are actually positive. Precision is defined as follows:

$P = precision = (TP)/(TP + FP)$ (3)

The fraction of positive tuples that were correctly classified as positive is recall:

$R = recall = (TP)/(TP + FN)$ (4)

The Receiver Operating Characteristic (ROC) analysis is done. The classification threshold from the most positive classification value to the most negative points of the ROC curve are obtained by sweeping. The area under the ROC curve (AUC)is called a quantitative summary of the ROC curve.

By the F-measure classification is also quantified, defined as the weighted harmonic mean of its precision and recall:

$F = 2PR/(P+R)$ (5)

The F-measure assumes values in the interval [0,1]. when no relevant instances have been retrievedit is 0, and if all retrieved instances are relevant is 1and all relevant instances have been retrieved. Experimental results are given in the following table 1,2 and 3.

Table 1 Using number of colors, texture features and morphology features classification results are produced

Classifier	**Accuracy**	**ROC**	**Average Precision**	**F-measure**
Multi-Layer Perceptron	0.641	0.523	0.638	0.6331
Support vector machine	0.79	0.6	0.596	0.672
Decision trees	0.718	0.418	0.634	0.658
KNN;k=3	0.725	0.430	0.62	0.653
Adaboost	0.731	0.499	0.725	0.67
Bagging	0.732	0.53	0.628	0.68
Stacking	0.78	0.474	0.629	0.672
MAEoC	0.888	0.789	0.903	0.876
MAE-NBC	0.890	0.790	0.913	0.887

Table 2

Using number of colors, texture features, morphology features, and RGB values obtained in the color segmentation phase classification results are produced

Classifier	**Accuracy**	**ROC**	**Average Precision**	**F-measure**
Multi-Layer Perceptron	0.676	0.493	0.683	0.691
Support vector machine	0.79	0.6	0.596	0.672
Decision trees	0.678	0.478	0.614	0.629
KNN;k=3	0.736	0.490	0.639	0.673
Adaboost	0.731	0.407	0.582	0.637
Bagging	0.76	0.509	0.589	0.659

Stacking	0.78	0.474	0.595	0.672
MAEoC	0.88	0.729	0.864	0.806
MAE-NBC	0.89	0.740	0.913	0.887

Table 3

Using number of colors, texture features, morphology features, RGB values obtained in the color segmentation phase, and the area of the lesion classification results are produced

Classifier	**Accuracy**	**ROC**	**Average Precision**	**F-measure**
Multi-Layer Perceptron	0.69	0.493	0.638	0.671
Support vector machine	0.79	0.6	0.596	0.672
Decision trees	0.632	0.478	0.644	0.658
KNN;k=3	0.736	0.490	0.639	0.673
Adaboost	0.711	0.407	0.725	0.67
Bagging	0.746	0.519	0.628	0.678
Stacking	0.78	0.474	0.629	0.672
MAEoC	0.88	0.669	0.903	0.806
MAE-NBC	0.89	0.670	0.913	0.887

Comparison of existing system with proposed system

Author	**Dataset**	**Pre-processing**	**Feature extraction method**	**Classifier**	**Detection performance**
Proposed system	ISIC repository	K-Means Clustering	GLSM	Ensemble of classifiers	Accuracy=96.20%
Heydy Castillejos-Fernández[1]	ISIC repository	Arti craft removal with razor algorithm	Fuzzy discrete wavelet transformation	Multi agent ensemble of classifiers	Accuracy=88%
Immagulate & Vijaya [2]	Dermnet/Dermofit	Image resizing	Colour and texture features	Support vector machine	Accuracy=86%
Mengistu [3]	Dermquest/Dermnet	Median Filtering	GLCM and colour features	Self-organizing maps andradial basics functions	Accuracy=96.15%
Elgamal [4]	From a digital camera	Gaussian-median filter	Principal component analysis	Artificial neural network	Accuracy=95%
Sheha et AI [5]	Dermoscopy atlases	Resizing and colour space transformation	GLCM	Multi-layer perceptron	Accuracy=86%

Comparing to all other existing systems, the proposed system produces the higher efficiency of 96.20%.

III. CONCLUSION

From this research work, skin cancer images are segmented, feature and color are extracted. Finally learning and prediction is done to produce output. The aim is to help the patients and doctors to identify the skin cancer without going to hospitals. This diagnosis research work includes segmentation with k-Means clustering algorithm and feature extraction, color extraction using ensemble of cluster, creation of feature vector using GLSM, learning and prediction using ensemble of classifier with the highest accuracy of 96.20% which shows promising results.

REFERENCE

[1] Heydy Castillejos-Fernándeza, Omar López-Ortegaa, Félix Castro-Espinozaa and Volodymyr Ponomaryovb. An Intelligent System for the Diagnosis of Skin Cancer on Digital Images taken with Dermoscopy. Vol. 14, No. 3, 2017

[2] I. Immagulate and M. S. Vijaya. Categorization of Non-Melanoma Skin Lesion Diseases Using Support Vector Machine and Its Variants. International Journal of Medical Imaging. Vol. 3, No. 2, pp. 34-40, 2015

[3] A. D. Mengistu. Computer Vision for Skin Cancer Diagnosis and Recognition using RBF and SOM. International Journal of Image Processing. Vol. 9, pp. 311-319, 2015

[4] M. Elgamal. Automatic Skin Cancer Images Classification. International Journal of Advanced Computer Science and applications. Vol. 4, No. 3, pp. 287-294, 2013

[5] M. Sheha, M. Mabrouk and A. Sharawy. Automatic detection of melanoma skin cancer using texture analysis. International Journal of Computer Applications. Vol. 42, No. 20, pp. 22-26, 2012